What is population structure

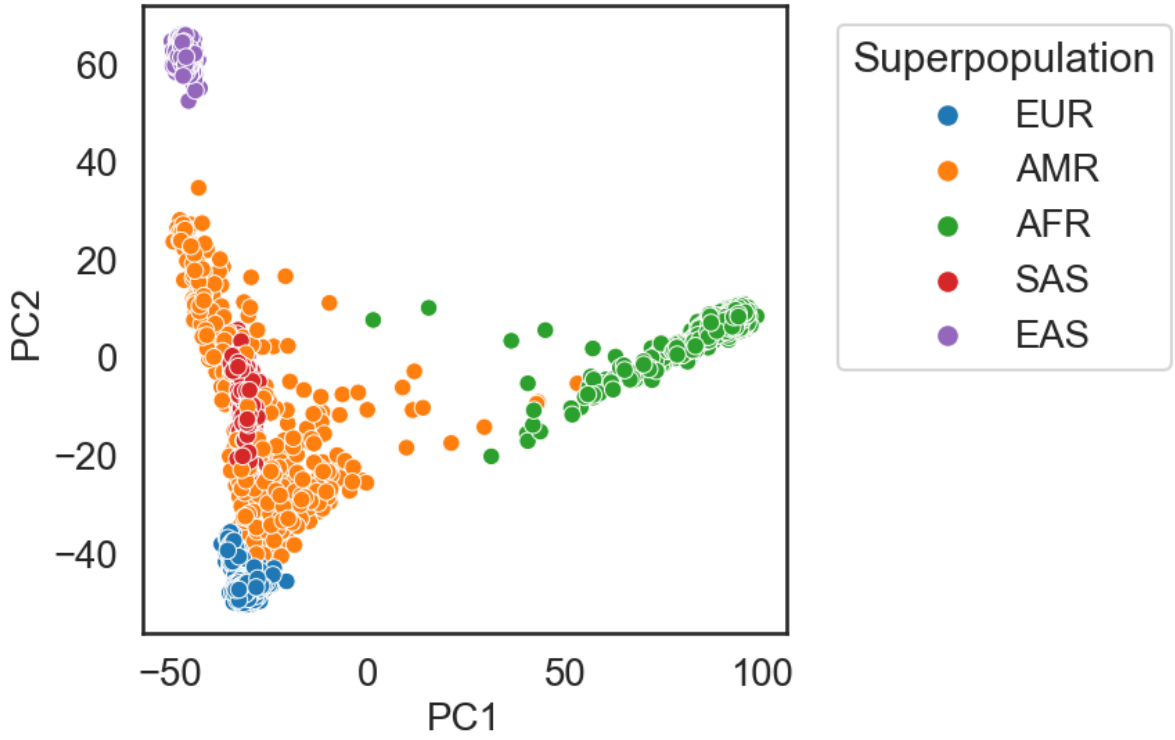
- Population genetic structure

- Individuals within a subgroup are more genetically similar to each other than to individuals in other subgroups.

- By extracting extensive variant information, such as SNPs and STRs, we can explore the genetic architecture of the population.

Methods of Population Structure Analysis

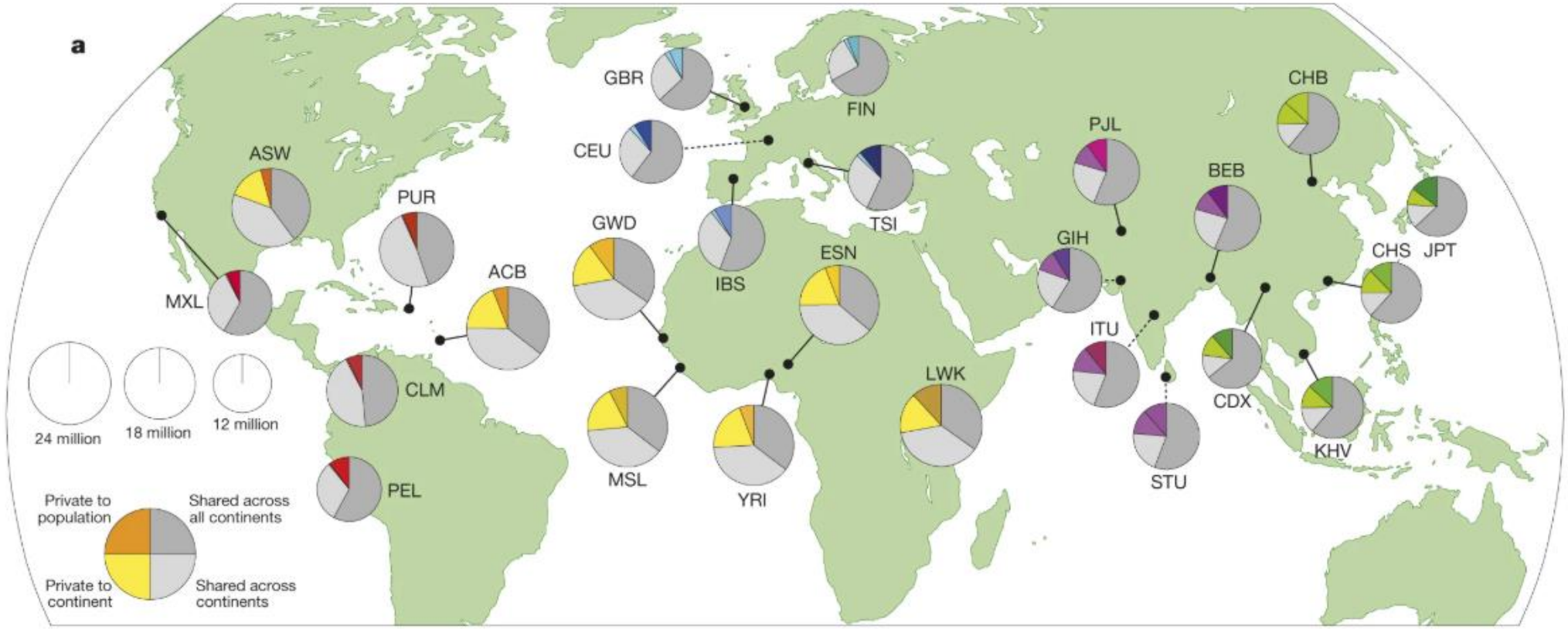
- Principal Component Analysis (PCA)



Methods of Population Structure Analysis

- ADMIXTURE

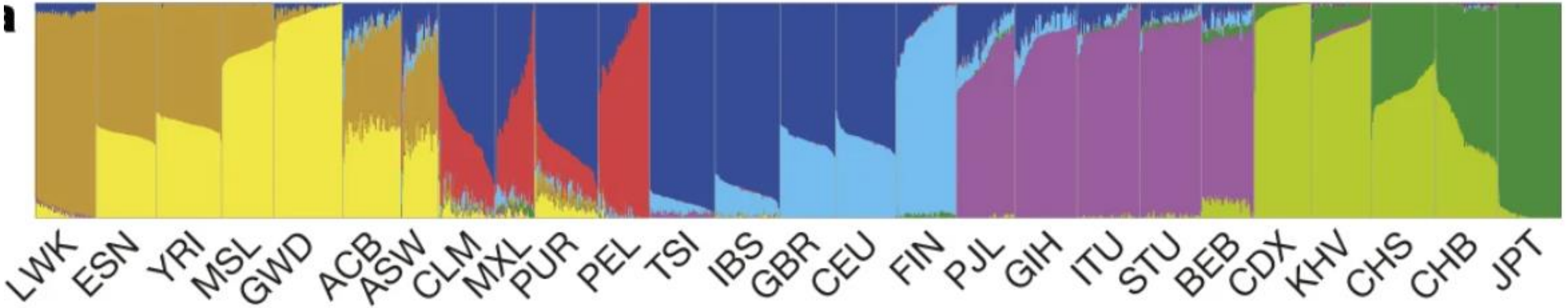
It models each individual as a mixture of K putative populations and estimate the proportions of individuals from each population.



The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015). https://doi.org/10.1038/nature15393

Methods of Population Structure Analysis

Population structure inferred using the ADMIXTURE program for K = 8 from1000 genomes project



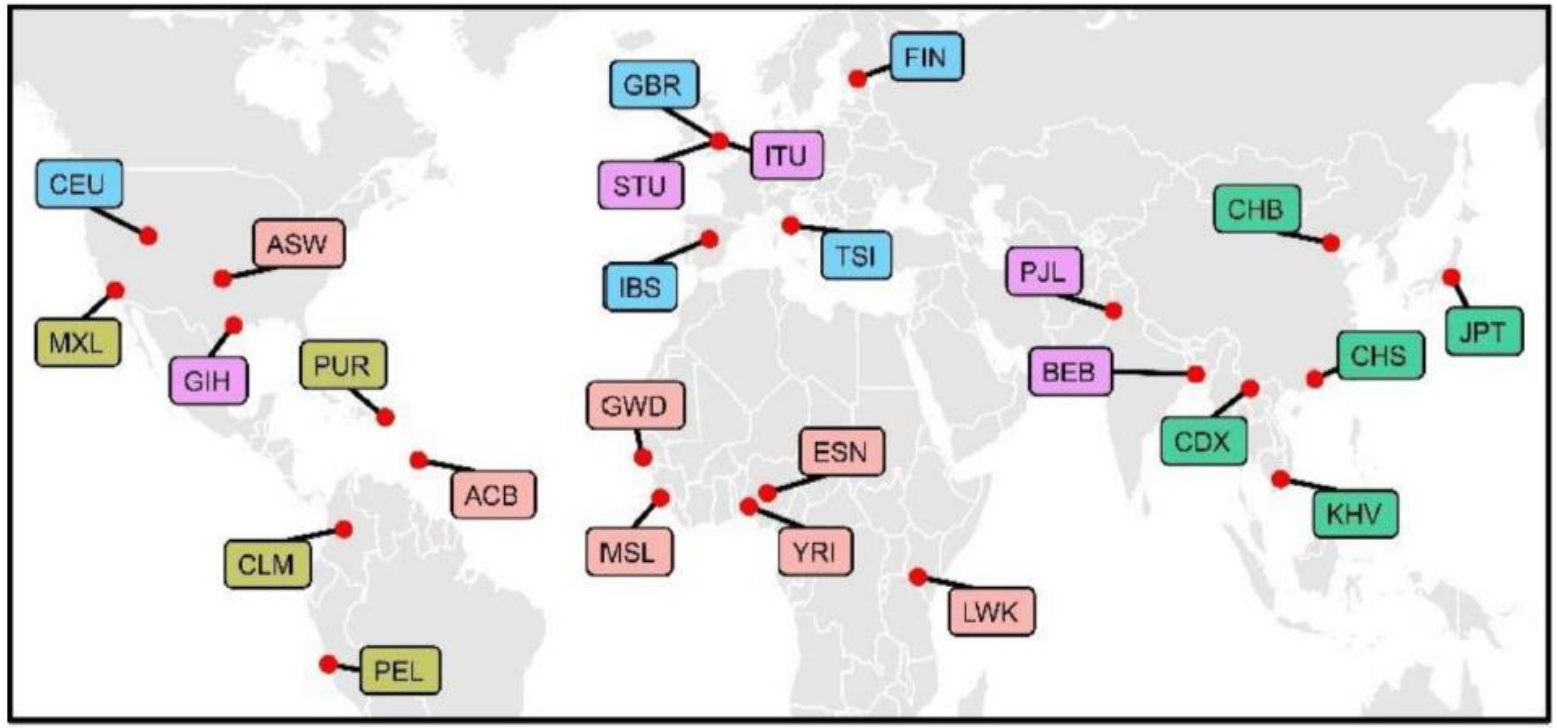
The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015). https://doi.org/10.1038/nature15393

Projects

Feifei Xia

30-04-2025

1000 Genome Projects



Superpopulation → Population

Superpopulation	Population
African Ancestry	ACB African Caribbean, Barbados
	ASW African American in Southwest US, US
	ESN Esan, Nigeria
	GWD Mandinka Gambian, The Gambia
	LWK Luhya in Webuye, Kenya
	MSL Mende, Sierra Leone
	YRI Yoruba in Ibadan, Nigeria
American Ancestry	CLM Colombian in Medellin, Colombia
	MXL Mexican Ancestry in California, US
	PEL Peruvian in Lima, Peru
	PUR Puerto Rican, US
East Asian Ancestry	CDX Chinese Dai in Xishuangbanna, China
	CHB Han Chinese in Beijing, China
	CHS Han Chinese South, China
	JPT Japanese in Tokyo, Japan
	KHV Kinh in Ho Chi Minh City, Vietnam
European Ancestry	CEU Northwest European Ancestry, US
	FIN Finnish, Finland
	GBR British, England and Scotland
	IBS Iberian, Spain
	TSI Toscani, Italy
Southeast Asian Ancestry	BEB Bengali, Bangladesh
	GIH Gujarati Indians, TX, US
	ITU Indian Telugu, UK
	PJL Punjabi in Lahore, Pakistan
	STU Sri Lankan Tamil, UK

Project 1. Population structure analysis using SNPs

Dataset

https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/

The SNV genotypes are made available in VCF files for each chromosome. You can start with SNV on chr1.

Workflow

VCF file -- PLINK convert -- LD Pruning -- PCA analysis -- ADMIXTURE Analysis

Reading

https://link.springer.com/protocol/10.1007/978-1-0716-0199-0_4

https://connor-french.github.io/intro-pop-structure-r/

Project 2. Population structure analysis using STRs

Dataset

https://drive.google.com/drive/folders/1fEy09eRa0Cs4O_paZvyO5rAwnfvdt7M-?usp=sharing

The STR genotypes are available in CSV files for each chromosome. You can start with STR on one chromosome.

Workflow

CSV file -- Filtering -- PCA analysis -- Clustering analysis – Supervised classification

Reading

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-024-05703-y

Project 3. Survival Analysis

Dataset

https://progenetix.org/subsets/NCIT-subsets/

Workflow

Select TCGA tumor type -- compare two survival model – Check in progenetix database

Reference

https://www.emilyzabor.com/survival-analysis-in-r.html

https://ramaanathan.github.io/SurvivalAnalysis/

https://bioconductor.org/packages/devel/bioc/vignettes/pgxRpi/inst/doc/Introduction_1_load_metadata.html

Presentations

- Define the project goal
- What kind of results do you want to show?
- Introduce the movitation, dataset, methods and present your results in 20 minutes